Linear Programming Relaxation Approach to Weighted CSP

Tomáš Werner

Center for Machine Perception
Czech Technical University
Prague, Czech Republic

Outline

- Notation
- Part 1: LP relaxation approach of WCSP, following [Schlesinger-1976]
- Part 2: Algorithms to solve the LP relaxation
- Part 3: Higher order relaxations

Notation

V	(finite) set of variables, totally ordered
$v \in V$	a single variable
X_v	(finite) domain of variable $v \in V$
$x_v \in X_v$	state of variable $v \in V$
$A \subseteq V$	a subset of variables
$X_A = \prod_{v \in A} X_v$	joint domain of variables $A \subseteq V$ (ordered by the order on V)
$x_A \in X_A$	joint state ('tuple', 'configuration') of variables $A \subseteq V$

Convention: "Implicit restriction"

For $B \subset A$, if symbols x_A and x_B appear in the same logical expression, x_B denotes the restriction of joint state x_A onto variables B.

$\bar{\mathbb{R}}$	extended reals, $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty\}$
$f_A \colon X_A \to \bar{\mathbb{R}}$	constraint with scope $A \subseteq V$
2^V	the set of all subsets of V
$\binom{V}{k}$	the set of all k-element subsets of V

Definition (Constraint network)

Let $E \subseteq 2^V$ be a hypergraph. Let each hyperedge $A \in E$ be assigned a constraint $f_A\colon X_A \to \bar{\mathbb{R}}$. This collection of constraints is called a constraint network.

Denoting $T(E) = \{\, (A, x_A) \mid A \in E,\ x_A \in X_A \,\}$, a constraint network is a mapping

$$\begin{aligned} f : \quad T(E) &\to \bar{\mathbb{R}} \\ (A, x_A) &\mapsto f_A(x_A) \end{aligned}$$

Definition (Weighted CSP, WCSP)

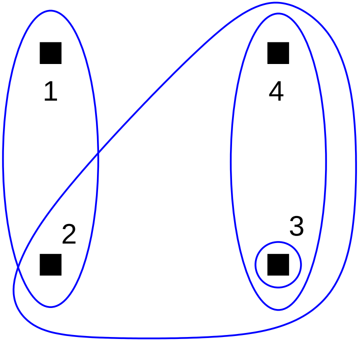
Given a constraint network f, compute the value (and, optionally, a maximiser) of

$$\max_{x_V \in X_V} \sum_{A \in E} f_A(x_A)$$

Example: A ternary WCSP

Let $V = (1, 2, 3, 4)$ and $E = \{(2, 3, 4), (1, 2), (3, 4), (3)\}$. Then

$$\max_{x_V} \sum_{A \in E} f_A(x_A) = \max_{x_1, x_2, x_3, x_4} [f_{234}(x_2, x_3, x_4) + f_{12}(x_1, x_2) + f_{34}(x_3, x_4) + f_3(x_3)]$$



Example: A binary WCSP

Let $E = \binom{V}{1} \cup E'$ with $E' \subseteq \binom{V}{2}$ (all unary and some binary constraints)

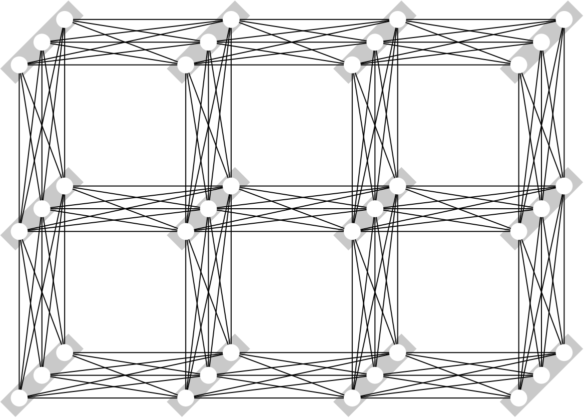
$$\max_{x_V} \sum_{A \in E} f_A(x_A) = \max_{x_V} \Big[\sum_{v \in V} f_v(x_v) + \sum_{vv' \in E'} f_{vv'}(x_v, x_{v'}) \Big]$$

Example: A binary WCSP

Let $E = \binom{V}{1} \cup E'$ with $E' \subseteq \binom{V}{2}$ (all unary and some binary constraints)

$$\max_{x_V} \sum_{A \in E} f_A(x_A) = \max_{x_V} \Big[\sum_{v \in V} f_v(x_v) + \sum_{vv' \in E'} f_{vv'}(x_v, x_{v'}) \Big]$$

Microstructure for E a grid graph and $X_v = \{1, 2, 3\}$:

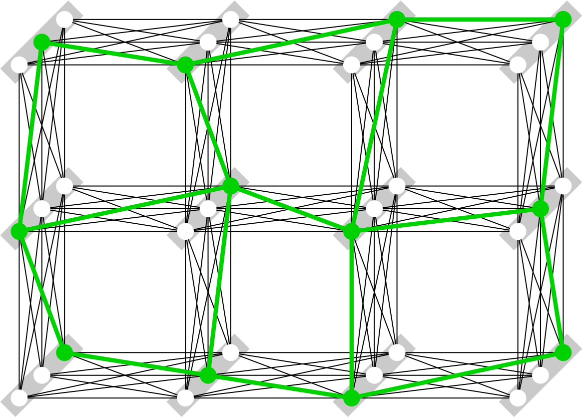


Example: A binary WCSP

Let $E = \binom{V}{1} \cup E'$ with $E' \subseteq \binom{V}{2}$ (all unary and some binary constraints)

$$\max_{x_V} \sum_{A \in E} f_A(x_A) = \max_{x_V} \Big[\sum_{v \in V} f_v(x_v) + \sum_{vv' \in E'} f_{vv'}(x_v, x_{v'}) \Big]$$

Microstructure for E a grid graph and $X_v = \{1, 2, 3\}$:



Part 1

LP Relaxation of WCSP by [Schlesinger-1976]
(more precisely, its n-ary generalisation)

Outline of Part 1

- Formulate two LPs, each yielding an upper bound on WCSP:
  - The first LP is a continuous relaxation of an integer LP formulation of WCSP
  - The second LP minimises an upper bound on WCSP by equivalent transformations
- Show that the two LPs are dual to each other
- Characterise when the upper bound is tight or minimal

Note: LP relaxation of WCSP yielding the same bound as [Schlesinger-1976] was proposed also by other researchers [Koster-1998, Chekuri-2001, Wainwright-2003, Cooper-deGivry-Schiex-2007]. We follow [Schlesinger-76] for its particular simplicity.

Primal LP

Joint states as probability distributions

Each hyperedge $A \in E$ is assigned a function $\mu_A \colon X_A \to \mathbb{R}$ that satisfies:

$$\begin{array}{lll} \text{(non-negativity)} & \mu_A(x_A) \geq 0 & A \in E;\ x_A \in X_A \\ \text{(normalisation)} & \displaystyle\sum_{x_A} \mu_A(x_A) = 1 & A \in E \end{array}$$

- Non-negativity and normalisation impose that μ_A is a probability distribution.

Each hyperedge $A \in E$ is assigned a function $\mu_A \colon X_A \to \mathbb{R}$ that satisfies:

$$
\begin{array}{lll}
\text{(non-negativity)} & \mu_A(x_A) \geq 0 & A \in E;\ x_A \in X_A \\
\text{(normalisation)} & \displaystyle\sum_{x_A} \mu_A(x_A) = 1 & A \in E \\
\text{(marginalisation)} & \displaystyle\sum_{x_{A \setminus B}} \mu_A(x_A) = \mu_B(x_B) & A, B \in E;\ B \subset A;\ x_B \in X_B
\end{array}
$$

- Non-negativity and normalisation impose that μ_A is a probability distribution.
- Marginalisation imposes that μ_B is the marginal distribution of μ_A.

Each hyperedge $A \in E$ is assigned a function $\mu_A \colon X_A \to \mathbb{R}$ that satisfies:

$$\begin{aligned}
&\text{(non-negativity)} && \mu_A(x_A) \geq 0 && A \in E;\ x_A \in X_A \\
&\text{(normalisation)} && \sum_{x_A} \mu_A(x_A) = 1 && A \in E \\
&\text{(marginalisation)} && \sum_{x_{A \setminus B}} \mu_A(x_A) = \mu_B(x_B) && A, B \in E;\ B \subset A;\ x_B \in X_B
\end{aligned}$$

- Non-negativity and normalisation impose that μ_A is a probability distribution.
- Marginalisation imposes that μ_B is the marginal distribution of μ_A.

Example

Let $A = (1, 2, 3, 4)$ and $B = (1, 3)$. The marginalisation condition reads

$$\mu_{13}(x_1, x_3) = \sum_{x_2, x_4} \mu_{1234}(x_1, x_2, x_3, x_4)$$

Each hyperedge $A \in E$ is assigned a function $\mu_A \colon X_A \to \mathbb{R}$ that satisfies:

$$\begin{array}{lll} \text{(non-negativity)} & \mu_A(x_A) \geq 0 & A \in E;\ x_A \in X_A \\ \text{(normalisation)} & \sum_{x_A} \mu_A(x_A) = 1 & A \in E \\ \text{(marginalisation)} & \sum_{x_{A \setminus B}} \mu_A(x_A) = \mu_B(x_B) & A, B \in E;\ B \subset A;\ x_B \in X_B \end{array}$$

- Non-negativity and normalisation impose that μ_A is a probability distribution.
- Marginalisation imposes that μ_B is the marginal distribution of μ_A.

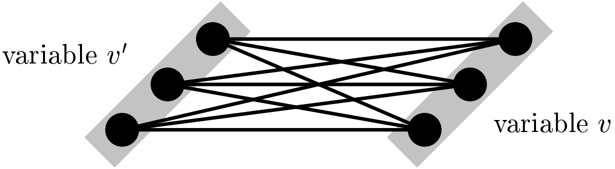
Example

Let $A = (1, 2, 3, 4)$ and $B = (1, 3)$. The marginalisation condition reads

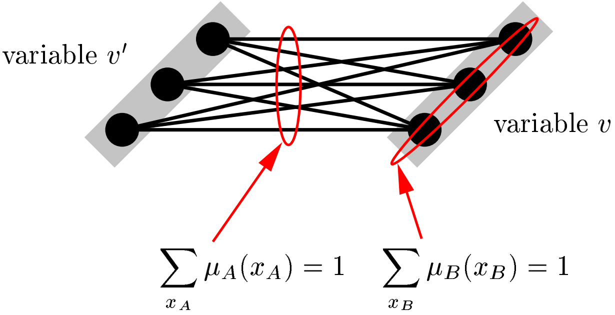
$$\mu_{13}(x_1, x_3) = \sum_{x_2, x_4} \mu_{1234}(x_1, x_2, x_3, x_4)$$

Note: To couple all pairs of overlapping distributions, we assume that E is closed under hyperedge intersection ($A, B \in E$ implies $A \cap B \in E$).

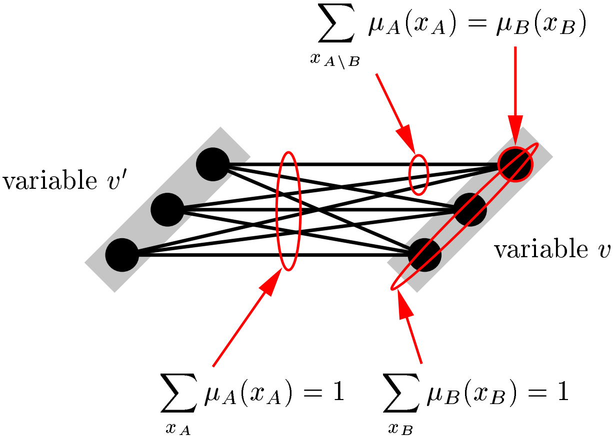
Example for $A = (v, v')$ and $B = (v)$:



Example for $A = (v, v')$ and $B = (v)$:



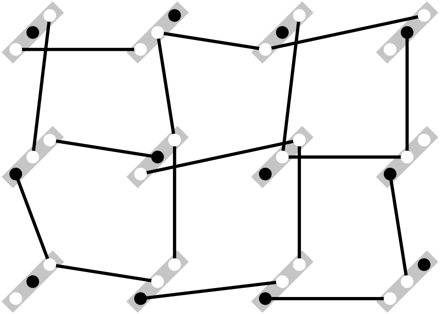
Example for $A = (v, v')$ and $B = (v)$:



What happens if the distributions μ are crisp, i.e., $\mu_A(x_A) \in \{0, 1\}$?

What happens if the distributions μ are crisp, i.e., $\mu_A(x_A) \in \{0, 1\}$?

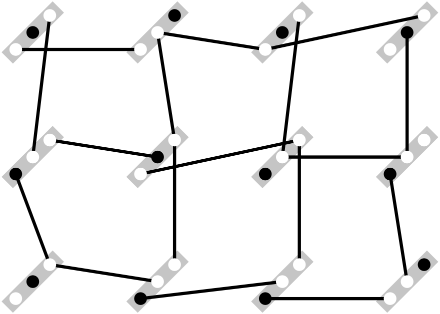
- Every distribution μ_A represents a single joint state.



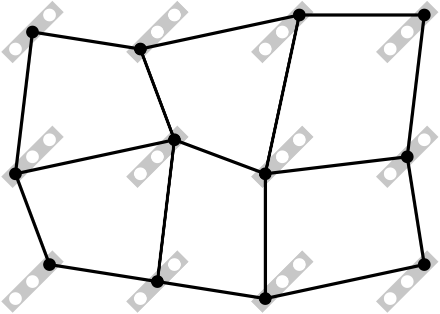
crisp μ satisfying normalisation

What happens if the distributions μ are crisp, i.e., $\mu_A(x_A) \in \{0, 1\}$?

- Every distribution μ_A represents a single joint state.
- The marginalisation constraint $\mu_B(x_B) = \sum_{x_{A \setminus B}} \mu_A(x_A)$ represents the fact that joint state μ_B is the restriction of joint state μ_A to variables $B \subset A$.



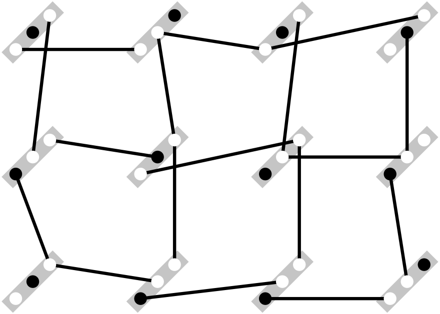
crisp μ satisfying normalisation



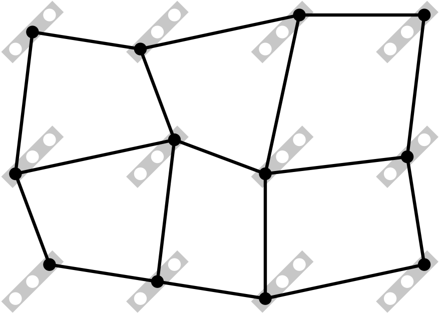
crisp μ satisfying also marginalisation

What happens if the distributions μ are crisp, i.e., $\mu_A(x_A) \in \{0, 1\}$?

- Every distribution μ_A represents a single joint state.
- The marginalisation constraint $\mu_B(x_B) = \sum_{x_{A \setminus B}} \mu_A(x_A)$ represents the fact that joint state μ_B is the restriction of joint state μ_A to variables $B \subset A$.
- The objective function is given by $f\,\mu = \sum_{A \in E} \sum_{x_A} f_A(x_A)\, \mu_A(x_A)$.



crisp μ satisfying normalisation



crisp μ satisfying also marginalisation

Theorem

Let $\binom{V}{1} \subseteq E$.
The WCSP optimum $\max\limits_{x_V} \sum\limits_{A \in E} f_A(x_A)$ *equals to the optimum of the integer LP*

$$
\begin{aligned}
& f\mu \to \max \\
\textit{subject to:} \quad & \mu_A(x_A) \in \{0,1\} && A \in E,\ x_A \in X_A \\
& \sum_{x_A} \mu_A(x_A) = 1 && A \in E \\
& \sum_{x_{A \setminus B}} \mu_A(x_A) = \mu_B(x_B) && A, B \in E,\ B \subset A,\ x_B \in X_B
\end{aligned}
$$

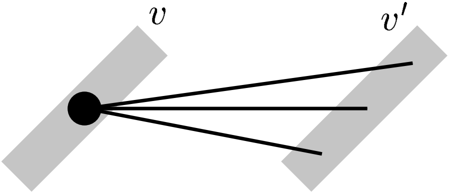
Relaxing $\mu_A(x_A) \in \{0,1\}$ to $\mu_A(x_A) \in [0,1]$ yields the primal LP.

Definition

- Networks f and f' are equivalent iff they yield the same objective function.
- A change of f to an equivalent network is an equivalent transformation.

An equivalent transformation is local iff it is applied to a triplet (A, B, x_B) with $B \subset A$ as follows:

Example for $A = (v, v')$ and $B = (v)$:



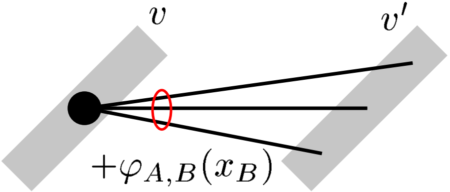
Definition

- Networks f and f' are equivalent iff they yield the same objective function.
- A change of f to an equivalent network is an equivalent transformation.

An equivalent transformation is local iff it is applied to a triplet (A, B, x_B) with $B \subset A$ as follows:

- add a constant $\varphi_{A,B}(x_B)$ to weights $\{ f_A(x_A) \mid x_{A \setminus B} \in X_{A \setminus B} \}$

Example for $A = (v, v')$ and $B = (v)$:



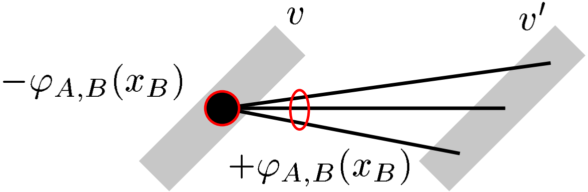
Definition

- Networks f and f' are equivalent iff they yield the same objective function.
- A change of f to an equivalent network is an equivalent transformation.

An equivalent transformation is local iff it is applied to a triplet (A, B, x_B) with $B \subset A$ as follows:

- add a constant $\varphi_{A,B}(x_B)$ to weights $\{ f_A(x_A) \mid x_{A \setminus B} \in X_{A \setminus B} \}$
- subtract the same constant from weight $f_B(x_B)$

Example for $A = (v, v')$ and $B = (v)$:



- Assign a constant $\varphi_{A,B}(x_B)$ to every such triplet (A, B, x_B) in the network. All these constants together are denoted by φ
- Applying local equivalent transformations on all triplets of a network f yields an equivalent network f^φ given by

$$f_A^\varphi(x_A) = f_A(x_A) + \sum_{B \subset A} \varphi_{A,B}(x_B) - \sum_{B \supset A} \varphi_{B,A}(x_A)$$

- Problems f^φ for all φ form an affine subspace of the space of all networks.
- For $f_A(x_A) > -\infty$, this subspace contains all networks equivalent with f.

Theorem (Upper bound on WCSP)

$$\underbrace{\max_{x_V} \sum_{A \in E} f_A(x_A)}_{WCSP\ optimum} \leq \sum_{A \in E} \max_{x_A} f_A(x_A)$$

The best upper bound is found by minimising it over equivalent networks:

$$\min_{\varphi} \sum_{A \in E} \max_{x_A} \Big[\underbrace{f_A(x_A) + \sum_{B \subset A} \varphi_{A,B}(x_B) - \sum_{B \supset A} \varphi_{B,A}(x_A)}_{f_A^{\varphi}(x_A)} \Big]$$

This can be written as a linear program

$$\begin{aligned} & \sum_{A \in E} \psi_A \to \min_{\varphi, \psi} \\ \text{subject to:} \quad & f_A^{\varphi}(x_A) \leq \psi_A \qquad A \in E,\ x_A \in X_A \end{aligned}$$

When is the upper bound tight?

Definition

Joint state x_A of hyperedge $A \in E$ is called $\begin{cases} \text{active} & \text{if} \quad f_A(x_A) = \max_{y_A} f_A(y_A) \\ \text{inactive} & \text{if} \quad f_A(x_A) < \max_{y_A} f_A(y_A) \end{cases}$

When is the upper bound tight?

Definition

Joint state x_A of hyperedge $A \in E$ is called $\begin{cases} \text{active} & \text{if} \quad f_A(x_A) = \max_{y_A} f_A(y_A) \\ \text{inactive} & \text{if} \quad f_A(x_A) < \max_{y_A} f_A(y_A) \end{cases}$

Theorem

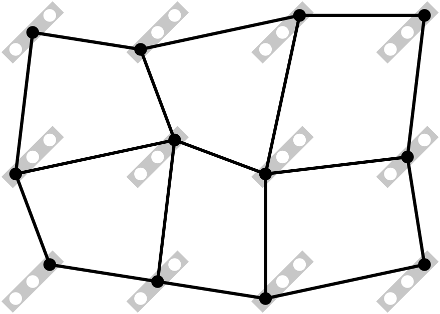
The upper bound is tight iff the (crisp) CSP formed by the active joint states is satisfiable.

When is the upper bound tight?

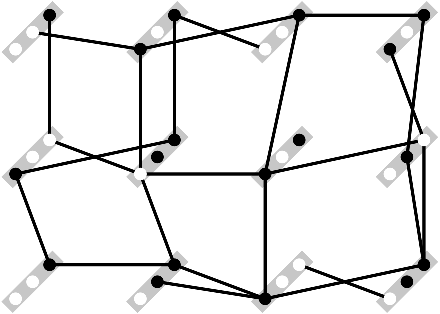
objective function for joint state x_V

upper bound

$$\sum_{A \in E} f_A(x_A) \quad \leq \quad \sum_{A \in E} \max_{x_A} f_A(x_A)$$



joint states given by x_V

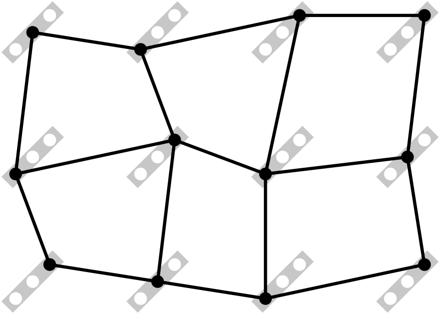


active joint states

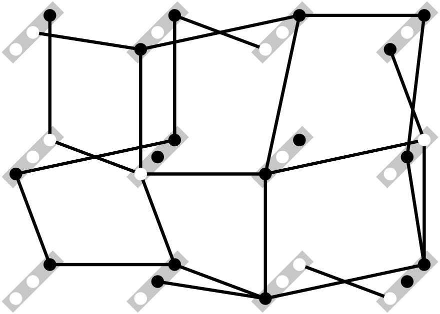
When is the upper bound tight?

objective function for joint state x_V | upper bound

$$\sum_{A \in E} f_A(x_A) \quad \leq \quad \sum_{A \in E} \max_{x_A} f_A(x_A)$$



joint states given by x_V



active joint states

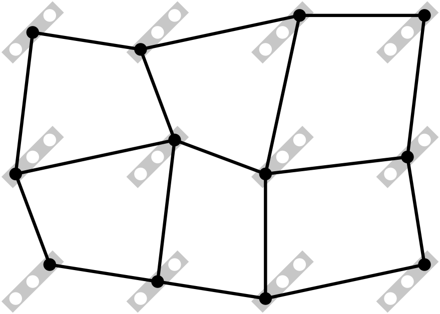
Do they form a satisfiable CSP?

When is the upper bound tight?

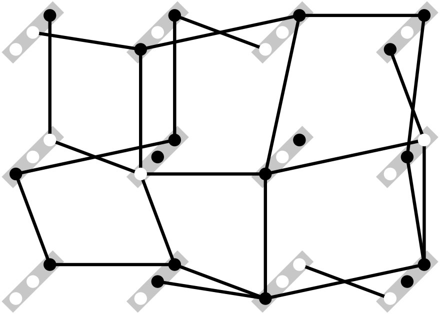
objective function for joint state x_V

upper bound

$$\sum_{A \in E} f_A(x_A) \quad < \quad \sum_{A \in E} \max_{x_A} f_A(x_A)$$



joint states given by x_V



active joint states

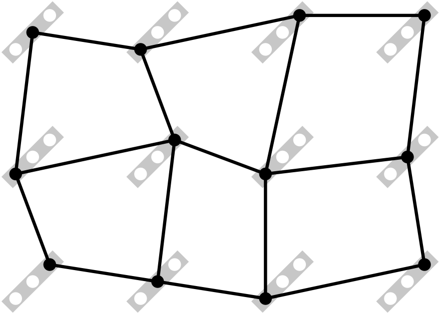
Do they form a satisfiable CSP?

No!

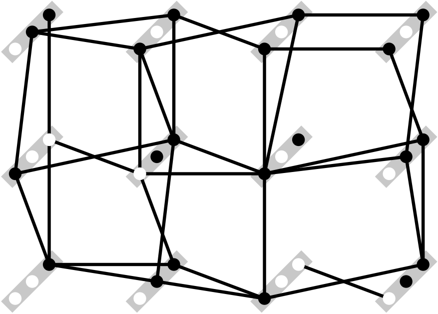
When is the upper bound tight?

objective function for joint state x_V upper bound

$$\sum_{A \in E} f_A(x_A) \quad \leq \quad \sum_{A \in E} \max_{x_A} f_A(x_A)$$



joint states given by x_V



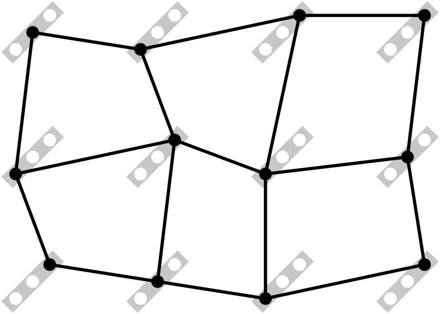
active joint states

Do they form a satisfiable CSP?

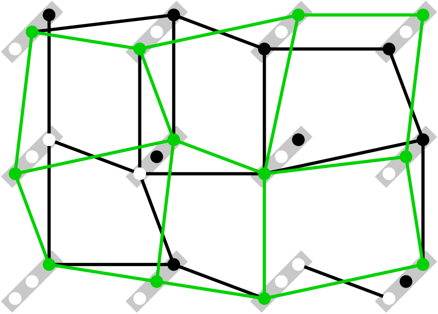
objective function for joint state x_V

upper bound

$$\sum_{A \in E} f_A(x_A) = \sum_{A \in E} \max_{x_A} f_A(x_A)$$



joint states given by x_V

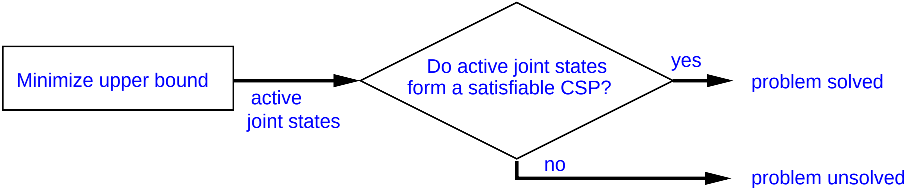


active joint states

Do they form a satisfiable CSP?

Yes!

The approach summarised



The instances that are solved exactly form a large, highly non-trivial tractable WCSP subclass.

LP duality

$$f\mu \to \max_{\mu} \qquad\qquad \sum_{A\in E} \psi_A \to \min_{\varphi,\psi}$$

$$\sum_{x_{A\setminus B}} \mu_A(x_A) = \mu_B(x_B) \qquad\qquad \varphi_{A,B}(x_B) \lessgtr 0 \qquad \begin{cases} A, B \in E \\ B \subset A \\ x_B \in X_B \end{cases}$$

$$\sum_{x_A} \mu_A(x_A) = 1 \qquad\qquad \psi_A \lessgtr 0 \qquad A \in E$$

$$\mu_A(x_A) \geq 0 \qquad \underbrace{f_A(x_A) + \sum_{B\subset A} \varphi_{A,B}(x_B) - \sum_{B\supset A} \varphi_{B,A}(x_A)}_{f_A^{\varphi}(x_A)} \leq \psi_A \qquad \begin{cases} A \in E \\ x_A \in X_A \end{cases}$$

$$f\mu \to \max_{\mu} \qquad \sum_{A \in E} \psi_A \to \min_{\varphi,\psi}$$

$$\sum_{x_{A \setminus B}} \mu_A(x_A) = \mu_B(x_B) \qquad \varphi_{A,B}(x_B) \lesseqgtr 0 \qquad \begin{cases} A, B \in E \\ B \subset A \\ x_B \in X_B \end{cases}$$

$$\sum_{x_A} \mu_A(x_A) = 1 \qquad \psi_A \lesseqgtr 0 \qquad A \in E$$

$$\mu_A(x_A) \geq 0 \qquad \underbrace{f_A(x_A) + \sum_{B \subset A} \varphi_{A,B}(x_B) - \sum_{B \supset A} \varphi_{B,A}(x_A)}_{f_A^{\varphi}(x_A)} \leq \psi_A \qquad \begin{cases} A \in E \\ x_A \in X_A \end{cases}$$

$$f\mu \to \max_{\mu} \qquad \psi 1 \to \min_{\varphi,\psi}$$

$$M\mu = 0 \qquad \varphi \lesseqgtr 0$$

$$N\mu = 1 \qquad \psi \lesseqgtr 0$$

$$\mu \geq 0 \qquad \varphi M + \psi N \geq f$$

LP duality

$$f\mu \to \max_{\mu} \qquad\qquad \sum_{A\in E}\psi_A \to \min_{\varphi,\psi}$$

$$\sum_{x_{A\setminus B}} \mu_A(x_A) = \mu_B(x_B) \qquad\qquad \varphi_{A,B}(x_B) \lessgtr 0 \qquad \begin{cases} A,B\in E \\ B\subset A \\ x_B\in X_B \end{cases}$$

$$\sum_{x_A} \mu_A(x_A) = 1 \qquad\qquad \psi_A \lessgtr 0 \qquad A\in E$$

$$\mu_A(x_A) \geq 0 \qquad\qquad \underbrace{f_A(x_A) + \sum_{B\subset A}\varphi_{A,B}(x_B) - \sum_{B\supset A}\varphi_{B,A}(x_A)}_{f_A^{\varphi}(x_A)} \leq \psi_A \qquad \begin{cases} A\in E \\ x_A\in X_A \end{cases}$$

$$f\mu \to \max_{\mu} \qquad\qquad \psi 1 \to \min_{\varphi,\psi}$$

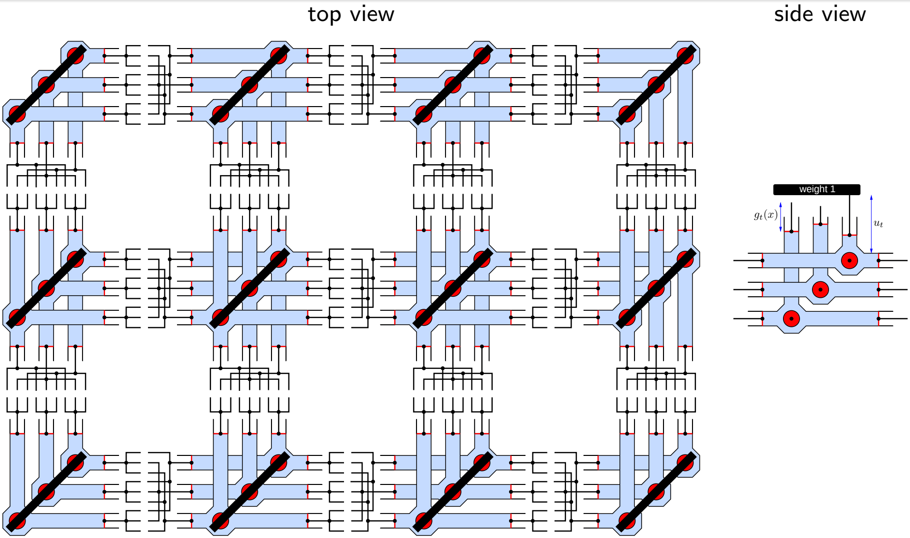
$$M\mu = 0 \qquad\qquad \varphi \lessgtr 0$$

$$N\mu = 1 \qquad\qquad \psi \lessgtr 0$$

$$\mu \geq 0 \qquad\qquad \varphi M + \psi N \geq f$$

primal program **dual program**

Hydraulic model of the LP pair [Schlesinger-Kovalevsky-1970's]



- $\mu_A(x_A)$ correspond to forces in forks and presures in tanks.
- $\varphi_{A,B}(x_A)$ correspond to displacements of the pistons.
- The upper bound corresnond to potential energy of the device.

LP duality theorems

$$\begin{aligned} f\mu &\to \max \\ M\mu &= 0 \\ N\mu &= 1 \\ \mu &\geq 0 \end{aligned} \qquad\qquad \begin{aligned} \psi 1 &\to \min \\ \varphi &\lesseqgtr 0 \\ \psi &\lesseqgtr 0 \\ \varphi M + \psi N &\geq f \end{aligned}$$

Feasible μ and (φ, ψ) satisfy:

weak duality: $f\mu \leq \psi 1$

strong duality: μ and (φ, ψ) are optimal iff $f\mu = \psi 1$

complementary slackness: μ and (φ, ψ) are optimal iff $(\varphi M + \psi N - f)\mu = 0$

Complementary slackness in WCSP

Expression $(\varphi M + \psi N - f)\mu = 0$ reads

$$\forall A \in E,\, x_A \in X_A : \quad \Big[\max_{y_A} f_A^\varphi(y_A) - f_A^\varphi(x_A)\Big]\mu_A(x_A) = 0$$

which means:

Theorem (Complementary slackness in WCSP)

Let μ be primal-feasible. The primal and dual LPs are simultaneously optimal iff every joint state x_A of every hyperedge $A \in E$ satisfies the implication

$$(A, x_A) \textit{ is inactive} \quad \Longrightarrow \quad \mu_A(x_A) = 0$$

Definition

The CSP formed by active joint states is relaxed-satisfiable iff there exists a feasible μ for which the above implication holds.

Theorem

A network f has the least upper bound iff the CSP formed by active joint states is relaxed-satisfiable.

A crisp CSP represented by $\sigma\colon T(E) \to \{0,1\}$

is relaxed satisfiable

iff there exists $\mu \le \sigma$ satisfying

$$\mu_A(x_A) \ge 0$$
$$\sum_{x_A} \mu_A(x_A) = 1$$
$$\sum_{x_{A\setminus B}} \mu_A(x_A) = \mu_B(x_B)$$

Three levels of consistency of crisp CSP

A crisp CSP represented by $\sigma\colon T(E) \to \{0,1\}$

is satisfiable

is relaxed satisfiable

iff there exists $\mu \leq \sigma$ satisfying

$$\mu_A(x_A) \in \{0,1\}$$
$$\sum_{x_A} \mu_A(x_A) = 1$$
$$\sum_{x_{A \setminus B}} \mu_A(x_A) = \mu_B(x_B)$$

$$\mu_A(x_A) \geq 0$$
$$\sum_{x_A} \mu_A(x_A) = 1$$
$$\sum_{x_{A \setminus B}} \mu_A(x_A) = \mu_B(x_B)$$

Three levels of consistency of crisp CSP

A crisp CSP represented by $\sigma\colon T(E) \to \{0,1\}$

is satisfiable	is relaxed satisfiable	has a non-empty pairwise consistency closure

iff there exists $\mu \le \sigma$ satisfying

$\mu_A(x_A) \in \{0,1\}$	$\mu_A(x_A) \ge 0$	$\mu_A(x_A) \in \{0,1\}$
$\sum_{x_A} \mu_A(x_A) = 1$	$\sum_{x_A} \mu_A(x_A) = 1$	$\max_{x_A} \mu_A(x_A) = 1$
$\sum_{x_{A \setminus B}} \mu_A(x_A) = \mu_B(x_B)$	$\sum_{x_{A \setminus B}} \mu_A(x_A) = \mu_B(x_B)$	$\max_{x_{A \setminus B}} \mu_A(x_A) = \mu_B(x_B)$

Three levels of consistency of crisp CSP

A crisp CSP represented by $\sigma\colon T(E) \to \{0,1\}$

is satisfiable $\implies$ is relaxed satisfiable $\implies$ has a non-empty pairwise consistency closure

iff there exists $\mu \leq \sigma$ satisfying

$$\mu_A(x_A) \in \{0,1\} \qquad \sum_{x_A} \mu_A(x_A) = 1 \qquad \sum_{x_{A\setminus B}} \mu_A(x_A) = \mu_B(x_B)$$

$$\mu_A(x_A) \geq 0 \qquad \sum_{x_A} \mu_A(x_A) = 1 \qquad \sum_{x_{A\setminus B}} \mu_A(x_A) = \mu_B(x_B)$$

$$\mu_A(x_A) \in \{0,1\} \qquad \max_{x_A} \mu_A(x_A) = 1 \qquad \max_{x_{A\setminus B}} \mu_A(x_A) = \mu_B(x_B)$$

A crisp CSP represented by $\sigma\colon T(E) \to \{0,1\}$

is satisfiable $\Longrightarrow$ is relaxed satisfiable $\Longrightarrow$ has a non-empty pairwise consistency closure

iff there exists $\mu \leq \sigma$ satisfying

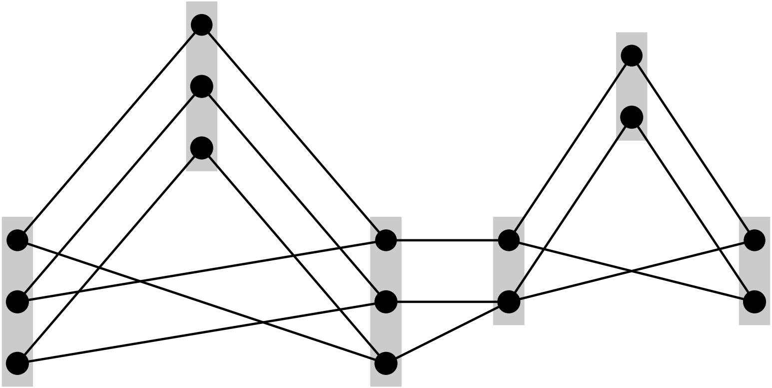
$$\mu_A(x_A) \in \{0,1\}$$
$$\sum_{x_A} \mu_A(x_A) = 1$$
$$\sum_{x_{A\setminus B}} \mu_A(x_A) = \mu_B(x_B)$$

$$\mu_A(x_A) \geq 0$$
$$\sum_{x_A} \mu_A(x_A) = 1$$
$$\sum_{x_{A\setminus B}} \mu_A(x_A) = \mu_B(x_B)$$

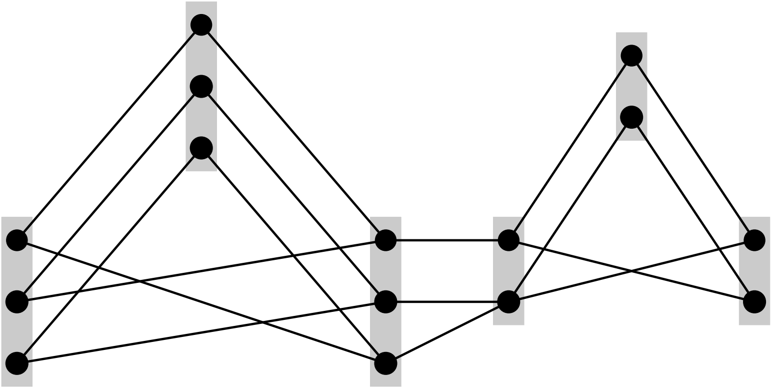
$$\mu_A(x_A) \in \{0,1\}$$
$$\max_{x_A} \mu_A(x_A) = 1$$
$$\max_{x_{A\setminus B}} \mu_A(x_A) = \mu_B(x_B)$$

upper bound is tight

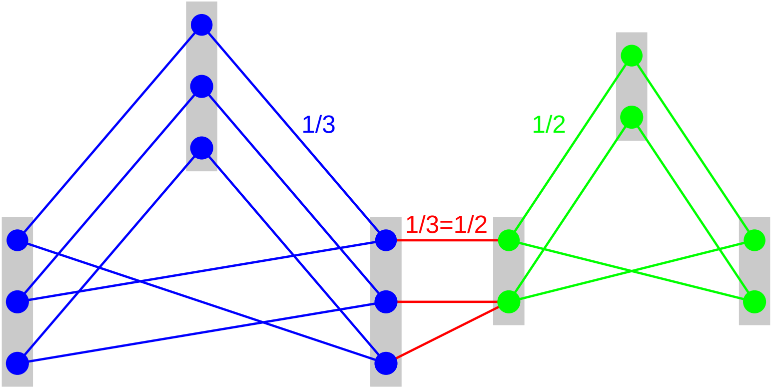
upper bound is minimal



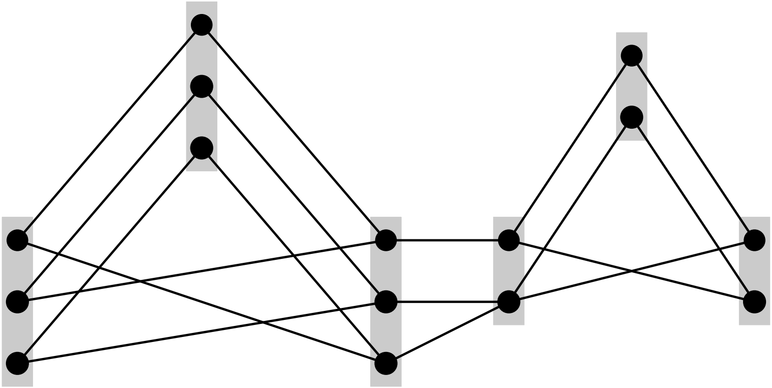
- This is an arc consistent CSP.



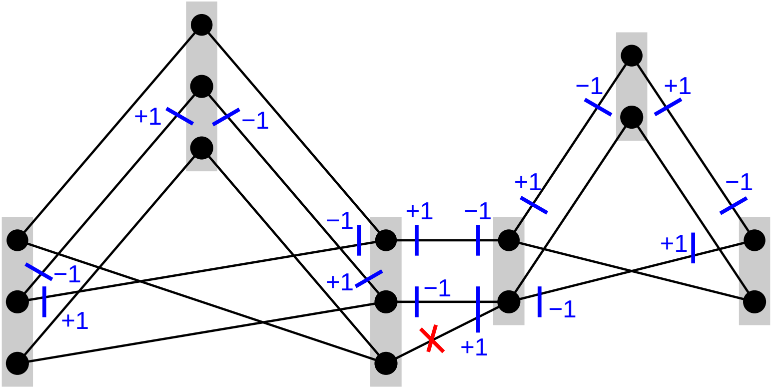
- This is an arc consistent CSP.
- But there exists no μ on allowed joint states satisfying non-negativity, normalisation and marginalisation condition.



- This is an arc consistent CSP.
- But there exists no μ on allowed joint states satisfying non-negativity, normalisation and marginalisation condition.



There exists an equivalent transformation of f that makes arc consistent closure empty. Hence, the upper bound can be decreased.



There exists an equivalent transformation of f that makes arc consistent closure empty. Hence, the upper bound can be decreased.

Part 2

Algorithms to Decrease the Upper Bound

The dual LP more suitable to solve than the primal LP because

- it has fewer variables,
- all the optimal solutions are encoded in active joint states.

We need to solve

$$\min_{\varphi} \sum_{A \in E} \max_{x_A} \Bigg[\underbrace{f_A(x_A) + \sum_{B \subset A} \varphi_{A,B}(x_B) - \sum_{B \supset A} \varphi_{B,A}(x_A)}_{f_A^{\varphi}(x_A)} \Bigg]$$

which is an unconstrained minimisation problem
with convex and nonsmooth (piecewise linear) objective function.

- Restriction on the algorithm: space complexity must be linear in the number of variables $\varphi_{A,B}(x_B)$.

- Algorithms that make active joint states pairwise/arc consistent
(more efficient, but not guaranteed to find global minimum of upper bound)

- Algorithms that find global minimum of the upper bound
(typically less efficient)

Survey of existing approaches

- Algorithms that make active joint states pairwise/arc consistent
  (more efficient, but not guaranteed to find global minimum of upper bound)
  - Max-sum diffusion [Koval-Kovalevsky-1976, Flach-2000, Werner-2008]
  - Sequential Tree-reweighted Message Passing (TRW-S) [Wainwright-etal-2003, Kolmogorov-2005]
  - Augmenting DAG algorithm [Koval-Schlesinger-1976]
    virtual arc consistency algorithm [Cooper-deGivry-Schiex-2006-8]

- Algorithms that find global minimum of the upper bound
  (typically less efficient)

- Algorithms that make active joint states pairwise/arc consistent
(more efficient, but not guaranteed to find global minimum of upper bound)

  - Max-sum diffusion [Koval-Kovalevsky-1976, Flach-2000, Werner-2008]
  - Sequential Tree-reweighted Message Passing (TRW-S) [Wainwright-etal-2003, Kolmogorov-2005]
  - Augmenting DAG algorithm [Koval-Schlesinger-1976]
virtual arc consistency algorithm [Cooper-deGivry-Schiex-2006-8]

  These algorithms yield exact solution for supermodular WCSPs!

- Algorithms that find global minimum of the upper bound
(typically less efficient)

- Algorithms that make active joint states pairwise/arc consistent
  (more efficient, but not guaranteed to find global minimum of upper bound)

  - Max-sum diffusion [Koval-Kovalevsky-1976, Flach-2000, Werner-2008]
  - Sequential Tree-reweighted Message Passing (TRW-S) [Wainwright-etal-2003, Kolmogorov-2005]
  - Augmenting DAG algorithm [Koval-Schlesinger-1976]
    virtual arc consistency algorithm [Cooper-deGivry-Schiex-2006-8]

  These algorithms yield exact solution for supermodular WCSPs!

- Algorithms that find global minimum of the upper bound
  (typically less efficient)

  - Subgradient descent [Schlesinger-Giginjak-2007, Komodakis-etal-2007]
  - Smoothing methods [Werner-2007, Johnson-etal-2007]

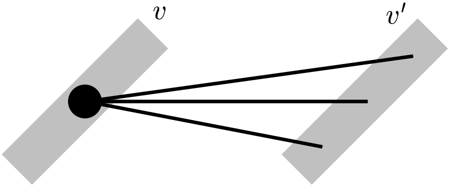
Max-sum diffusion repeats this simple operation:

- On a triplet (A, B, x_B) with $B \subset A$, apply the local equivalent transformation that makes satisfied the equality $\max_{x_{A \setminus B}} f_A(x_A) = f_B(x_B)$

Max-sum diffusion repeats this simple operation:

- On a triplet (A, B, x_B) with $B \subset A$, apply the local equivalent transformation that makes satisfied the equality $\max_{x_{A \setminus B}} f_A(x_A) = f_B(x_B)$

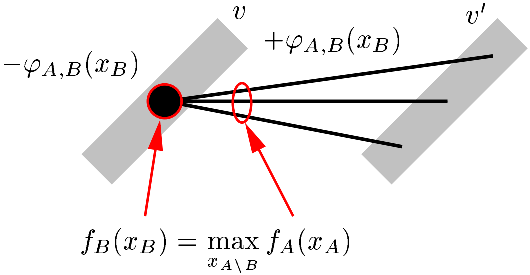
Example for $A = (v, v')$ and $B = (v)$:



Max-sum diffusion repeats this simple operation:

- On a triplet (A, B, x_B) with $B \subset A$, apply the local equivalent transformation that makes satisfied the equality $\max_{x_{A \setminus B}} f_A(x_A) = f_B(x_B)$

Example for $A = (v, v')$ and $B = (v)$:



Algorithm: Max-sum diffusion

```
loop
   for (A, B) such that A ∈ E, B ∈ E, B ⊂ A do
      for x_B ∈ X_B do
         φ_{A,B}(x_B) ← φ_{A,B}(x_B) + [f_B^φ(x_B) − max_{x_{A\B}} f_A^φ(x_A)]/2
      end for
   end for
end loop
```

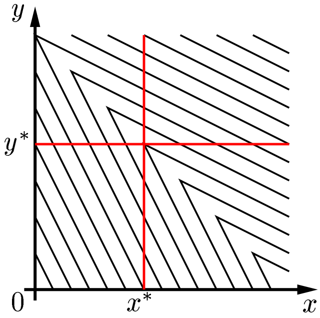
Properties of the algorithm:

- It monotonically decreases the upper bound.
- (Conjecture) It converges to a fixed point when $\max_{x_{A \setminus B}} f_A^{\varphi}(x_A) = f_B^{\varphi}(x_B)$ holds for all triplets (A, B, x_B) with $B \subset A$.
- At a fixed point, the CSP formed by active joint states is pairwise consistent.

Max-sum diffusion can be seen as a coordinate descent to decrease the upper bound:

- Minimise over a single free variable, keeping the other variables fixed.
- Iterate this for different free variables.

But coordinate descent is not guaranteed to find the global minimum of a convex nonsmooth function!



Point (x^*, y^*) is not a global minimum despite it is minimal separately in each coordinate.

Tree-reweighted Message Passing [Wainwright-etal-2003, Kolmogorov-2005]
Decomposing WCSP to tractable subproblems

- Let f^k denote a network with variables $V^k \subseteq V$ and hypergraph $E^k \subseteq 2^{V^k}$.
- Let $\{ f^k \mid k \in K \}$ be a collection of networks satisfying

$$f_A(x_A) = \sum_{k \mid A \in E^k} f_A^k(x_A) \qquad \forall A \in E,\ x_A \in X_A$$

- The objective function $F^k\colon X_{V^k} \to \bar{\mathbb{R}}$ of network f^k is $F^k(x_{V^k}) = \sum_{A \in E^k} f_A^k(x_A)$
- Swapping maximum and sum yields an upper bound on WCSP:

$$\max_{x_V} \sum_{A \in E} f_A(x_A) = \max_{x_V} \sum_k F^k(x_{V^k}) \le \sum_k \max_{x_{V^k}} F^k(x_{V^k})$$

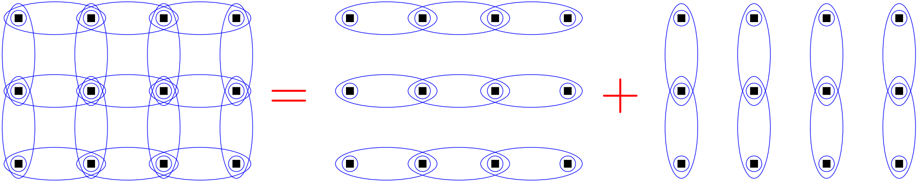
Minimising the upper bound

Minimise $\sum_k \max_{x_{V^k}} \sum_{A \in E^k} f_A^k(x_A)$ over collections $\{ f^k \mid k \in K \}$ subject to

$$f_A(x_A) = \sum_{k \mid A \in E^k} f_A^k(x_A)$$

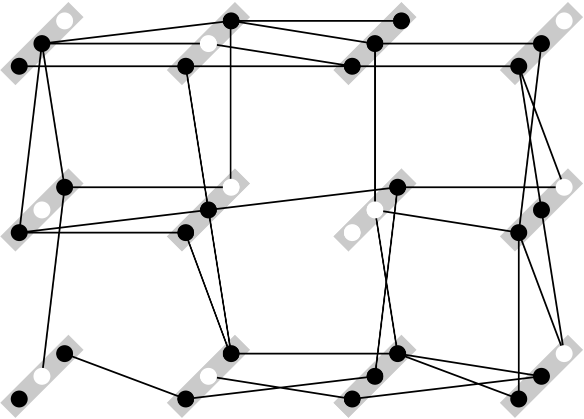
Tree-reweighted Message Passing [Wainwright-etal-2003, Kolmogorov-2005]

Example: Decomposing a grid graph to rows and columns

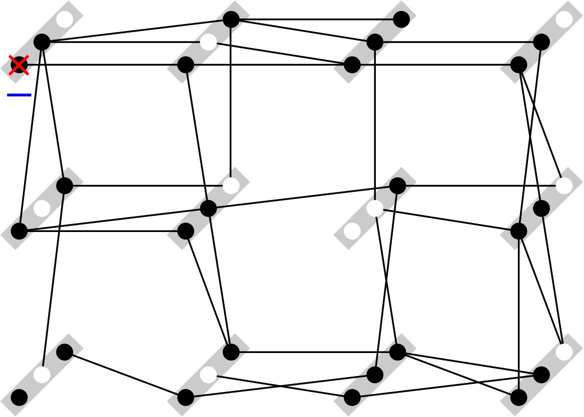


- Max-marginals of the row and column subproblems are iteratively equalised (like in max-sum diffusion) in the unary constraints.
- The row and column subproblems are solved efficiently by dynamic programming, in an incremental (hence efficient) way.

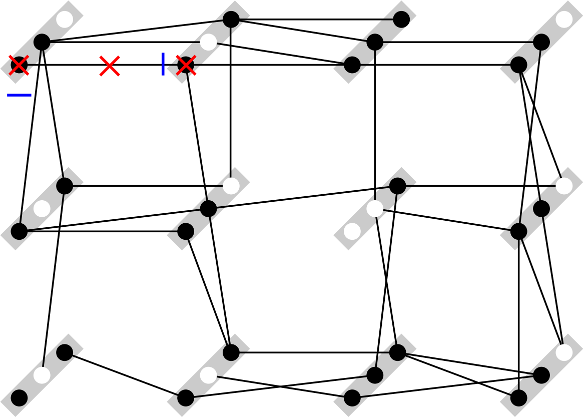
- Decomposing WCSP to subproblems is equivalent to approach by [Schlesinger-1976]. They can be translated to each other:
  - Expressing [Schlesinger-1976] by decomposition: Hypergraphs E^k are individual hyperedges $A \in E$.
  - Expressing decomposition by [Schlesinger-1976]: Each constraint f_A is itself a tractable WCSP, F^k.
- For binary problems, decomposing WCSP to trees yields (under mild assumptions) the same relaxation as the max-sum diffusion. But more efficient, especially on images.
- It is not straightforward to extend this efficiency to non-binary networks.



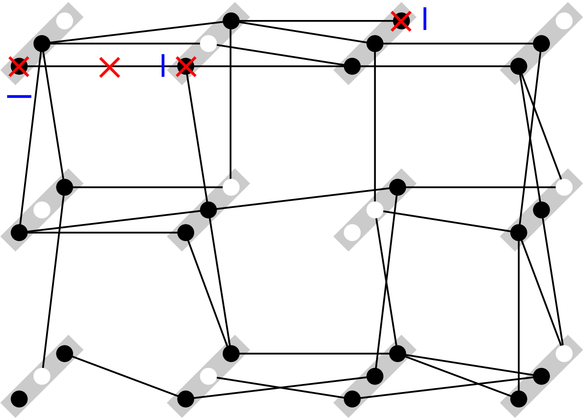
Run arc consistency algorithm on the active joint states. During this, remember pointers to causes of deletion. The pointers form an directed acyclic graph (DAG).



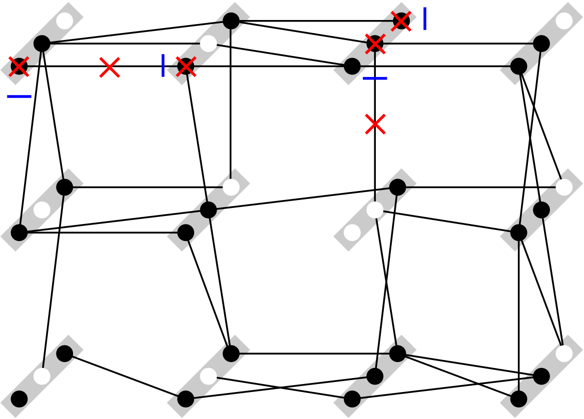
Run arc consistency algorithm on the active joint states. During this, remember pointers to causes of deletion. The pointers form an directed acyclic graph (DAG).



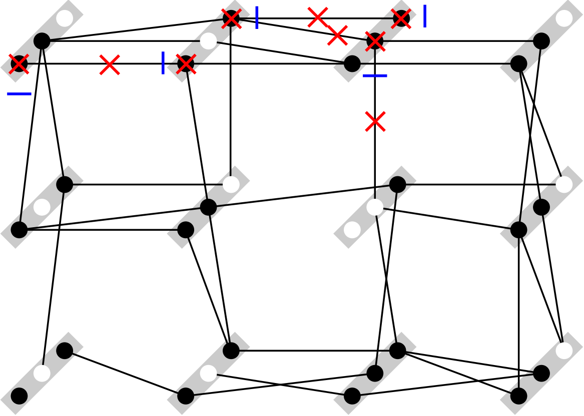
Run arc consistency algorithm on the active joint states. During this, remember pointers to causes of deletion. The pointers form an directed acyclic graph (DAG).



Run arc consistency algorithm on the active joint states. During this, remember pointers to causes of deletion. The pointers form an directed acyclic graph (DAG).

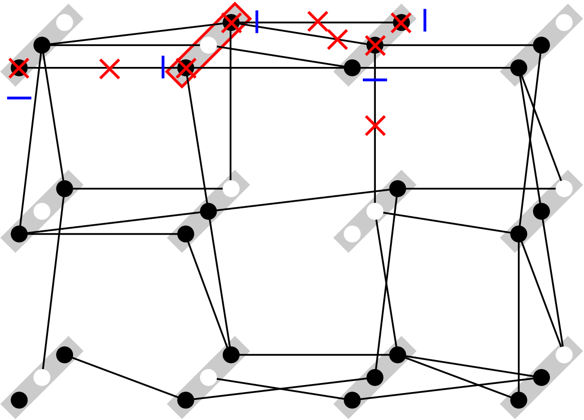


Run arc consistency algorithm on the active joint states. During this, remember pointers to causes of deletion. The pointers form an directed acyclic graph (DAG).



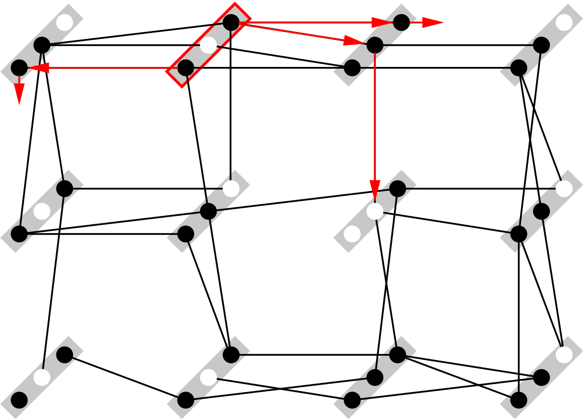
Run arc consistency algorithm on the active joint states. During this, remember pointers to causes of deletion. The pointers form an directed acyclic graph (DAG).

Augmenting DAG algorithm [Koval-Schlesinger-1976, Cooper-Schiex-2004]



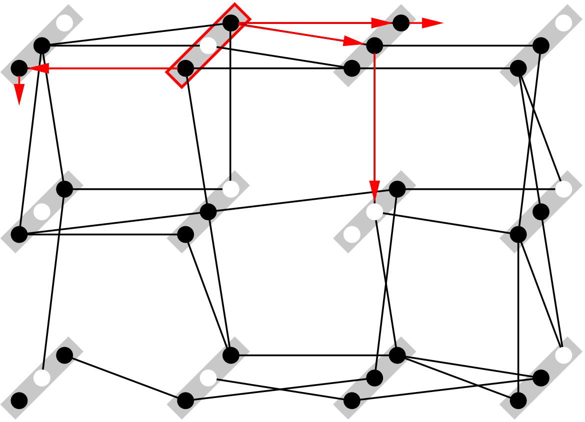
Run arc consistency algorithm on the active joint states. During this, remember pointers to causes of deletion. The pointers form an directed acyclic graph (DAG).

- If all states in any variable are deleted, the upper bound cannot be minimal.



Run arc consistency algorithm on the active joint states. During this, remember pointers to causes of deletion. The pointers form an directed acyclic graph (DAG).

- If all states in any variable are deleted, the upper bound cannot be minimal. Then, it can be decreased by pushing weights along (a subgraph of) the DAG.

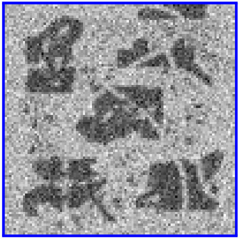


Run arc consistency algorithm on the active joint states. During this, remember pointers to causes of deletion. The pointers form an directed acyclic graph (DAG).

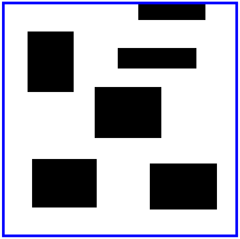
- If all states in any variable are deleted, the upper bound cannot be minimal. Then, it can be decreased by pushing weights along (a subgraph of) the DAG.
- If a non-empty arc consistency closure is found, halt.

Example WCSP: Syntactic image analysis

Find the image containing non-overlapping rectangles, nearest to input image!



input image

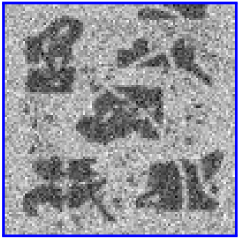


output image

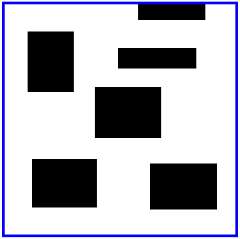
Example WCSP: Syntactic image analysis

Find the image containing non-overlapping rectangles, nearest to input image!

- Variables V are pixels, hypergraph E is the image grid.



input image

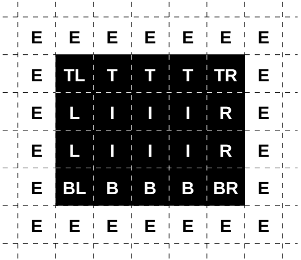


output image

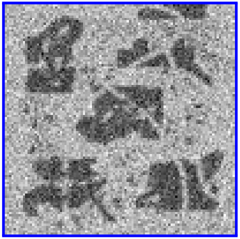
Example WCSP: Syntactic image analysis

Find the image containing non-overlapping rectangles, nearest to input image!

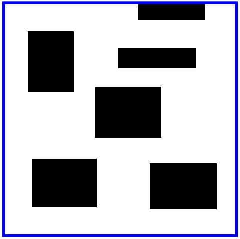
- Variables V are pixels, hypergraph E is the image grid.
- Variable domains $X_v = \{\,\mathrm{E}, \mathrm{I}, \mathrm{L}, \mathrm{R}, \mathrm{T}, \mathrm{B}, \mathrm{TL}, \mathrm{TR}, \mathrm{BL}, \mathrm{BR}\,\}$ are syntactic parts of a rectangle.



hidden states = syntactic parts



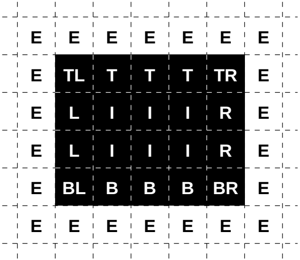
input image



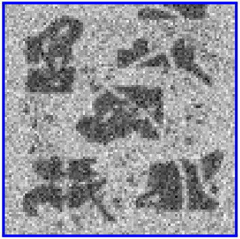
output image

Find the image containing non-overlapping rectangles, nearest to input image!

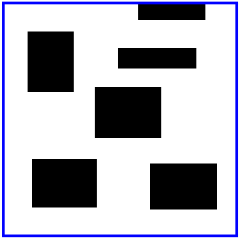
- Variables V are pixels, hypergraph E is the image grid.
- Variable domains $X_v = \{\,\mathrm{E},\ \mathrm{I},\ \mathrm{L},\ \mathrm{R},\ \mathrm{T},\ \mathrm{B},\ \mathrm{TL},\ \mathrm{TR},\ \mathrm{BL},\ \mathrm{BR}\,\}$ are syntactic parts of a rectangle.
- Unary constraint $f_v(x_v)$ quantifies agreement of intensity of state x_v and intensity of input pixel v.



hidden states = syntactic parts
observed states = {black,white}



input image



output image

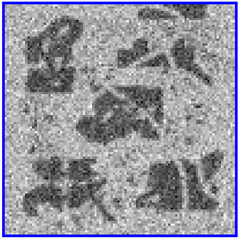
Example WCSP: Syntactic image analysis

Find the image containing non-overlapping rectangles, nearest to input image!

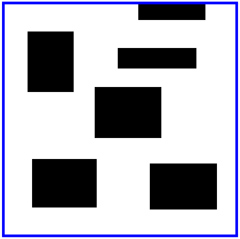
- Variables V are pixels, hypergraph E is the image grid.
- Variable domains $X_v = \{\, \text{E, I, L, R, T, B, TL, TR, BL, BR} \,\}$ are syntactic parts of a rectangle.
- Unary constraint $f_v(x_v)$ quantifies agreement of intensity of state x_v and intensity of input pixel v.
- Binary constraint $f_{vv'}(x_v, x_{v'})$ equals 0 if syntactic parts x_v and $x_{v'}$ are allowed to neighbor and $-\infty$ otherwise.

E	E	E	E	E	E	E
E	TL	T	T	T	TR	E
E	L	I	I	I	R	E
E	L	I	I	I	R	E
E	BL	B	B	B	BR	E
E	E	E	E	E	E	E

hidden states = syntactic parts
observed states = {black,white}



input image



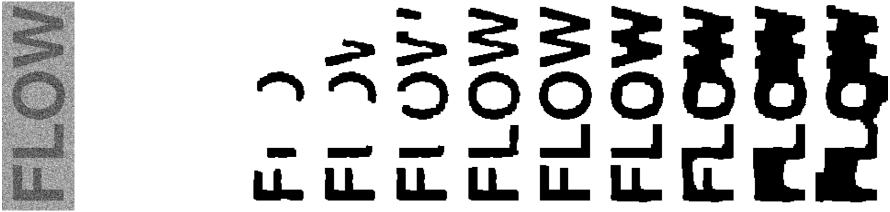
output image

Example: Binary WCSP with a global constraint [Werner-2008]

- $E = \binom{V}{1} \cup E' \cup \{V\}$ where $E' \subseteq \binom{V}{2}$, $X_v = \{\text{white}, \text{black}\}$.
- Unary constraint f_v quantifies agreement with intensity of input pixel v.
- Binary constraints $f_{vv'}$ penalise transition between black and white pixels.
- Global constraint $f_V(x_V)$ is: 0 if x_V contains n black pixels and $-\infty$ otherwise.

Interpretation: Find minimum *st*-cut in a graph such that the number of pixels in the first partition equals n (NP-hard).

- Max-sum diffusion enforces generalised arc consistency of active joint states.
- Equalising max-marginals between f_v and f_V seen as a soft global propagator.



input									
	n required:	2000	3000	4000	5000	5368	6000	7000	8000	9000
	n achieved:	2008	3004	4011	5006	5368	6004	7024	7982	9032

Supermodular problems

Let each domain X_v be totally ordered. A function f_A is supermodular if

$$f_A(x_A \wedge y_A) + f_A(x_A \vee y_A) \geq f_A(x_A) + f_A(y_A)$$

for any $x_A, y_A \in X_A$, where $\wedge$ ($\vee$) denotes the elementwise minimum (maximum).

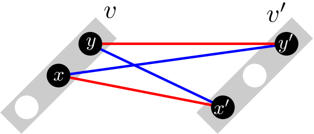
Theorem ([Schlesinger-Flach-00] for binary case, [Werner-2008] for non-binary case)

Let f_A be supermodular for each $A \in E$. Finding an equivalent network whose active joint states are generalised arc consistent solves the WCSP f exactly.

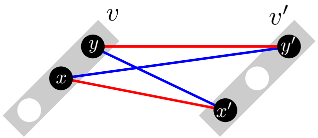
Proof

- Equivalent transformations preserve supermodularity.
- The maximisers of a supermodular function on a distributive lattice form a sublattice of this lattice [Topkis78]. Hence, the active joint states of each constraint form a lattice. Hence, all active joint states form a well-known tractable CSP [Jeavons-Cooper-95] (lattice CSP).
- Generalised arc consistency suffices for a lattice CSP to be satisfiable.

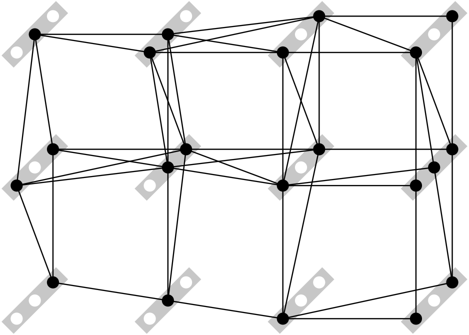
Before, [Cooper-2008] showed that the LP relaxation is tight for non-binary supermodular WCSPs. Our statement is stronger and the proof is simpler.

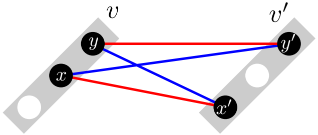


$$x \leq x', y \leq y' \quad \Longrightarrow \quad f_{vv'}(x, x') + f_{vv'}(y, y') \geq f_{vv'}(x, y') + f_{vv'}(y, x')$$

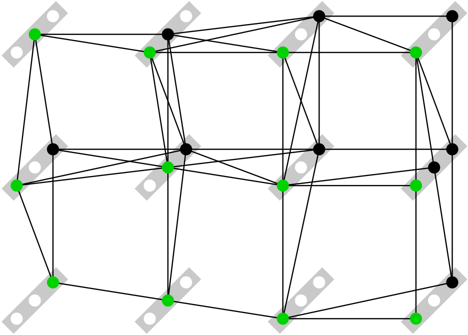


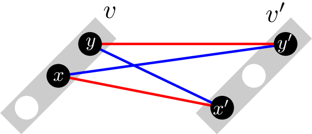
$$x \leq x', y \leq y' \quad \Longrightarrow \quad f_{vv'}(x, x') + f_{vv'}(y, y') \geq f_{vv'}(x, y') + f_{vv'}(y, x')$$



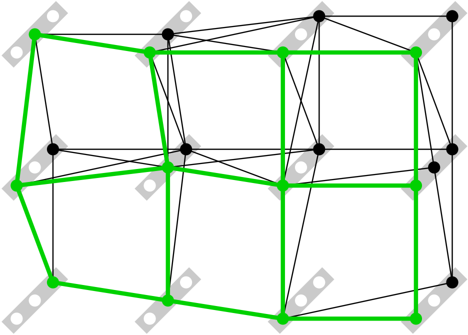


$$x \le x', y \le y' \quad \Longrightarrow \quad f_{vv'}(x, x') + f_{vv'}(y, y') \ge f_{vv'}(x, y') + f_{vv'}(y, x')$$





$$x \leq x', y \leq y' \quad \Longrightarrow \quad f_{vv'}(x, x') + f_{vv'}(y, y') \geq f_{vv'}(x, y') + f_{vv'}(y, x')$$



Subgradient descent

Well-known approach to minimise nonsmooth convex functions [Shor-1979].

Definition (Subgradient)

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a convex function. A vector $g(x_0) \in \mathbb{R}^n$ is a subgradient of f at point x_0 iff $g^\top (x - x_0) \le f(x) - f(x_0)$ for all $x \in \mathbb{R}^n$.

Theorem (Subgradient descent)

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a convex function. Let $\{\alpha_i\}_{i=0}^{\infty}$ be a sequence of positive numbers that converge to zero but their partial sums converge to infinity. For any initial point x_0, the sequence

$$x_{i+1} = x_i - \alpha_i \, g(x_i)$$

converges to $\min_{x \in \mathbb{R}^n} f(x)$.

Apply gradient descent to minimising the WCSP upper bound [Schlesinger-Giginjak-2007, Komodakis-etal-2007]:

- Converges (non-monotonically) to the global minimum of the upper bound.
- Inefficient if applied to a decomposition to individual hyperedges.
- Efficient if applied to a decomposition to longer trees/chains.

Smoothing methods

- Let $f_\beta\colon \mathbb{R}^n \to \mathbb{R}$ be a smooth convex function for every $\beta > 0$.
- With increasing β, let f_β approach a nonsmooth convex function f_∞.
- The sequence $\min_{x\in\mathbb{R}^n} f_\beta(x)$ converges to $\min_{x\in\mathbb{R}^n} f_\infty(x)$.

$\beta \leftarrow 1$
loop
    Minimise f_β by coordinate descent.
    Increase β.
end loop

An example [Werner-2007, Johnson-etal-2007]:

- Approach crisp maximum by a sequence of soft maxima:

$$\max\{x, y\} = \lim_{\beta\to\infty} \frac{1}{\beta} \log\left(e^{\beta x} + e^{\beta y}\right)$$

  Minimising the smoothed upper bound leads to geometric programming.
- Not very practical (in my experience) because of the two nested loops in the algorithm.

Part 3

Higher Order Polyhedral Relaxations

Adding zero constraints

Let f be a network with hypergraph E. Let $A \notin E$. Add constraint $f_A \equiv 0$ to f.

Theorem

By adding a zero constraint to a network, the optimal value of the LP pair

- *never increases*
- *for some instances, strictly decreases.*

Proof, by primal argument

- The primal objective function is preserved.
- The new primal feasible set is a strict subset of the old one.

Proof, by dual argument

- The dual objective function (i.e., the upper bound) is preserved.
- New equivalent transformations are enabled and no existing ones disabled. Hence, the new dual feasible set is a strict superset of the old one.

The proofs are imprecise because the dimensions of the old and new feasible sets are different. Let us be more precise...

primal program	**dual program**	
$f\mu \to \max$	$\sum_{A\in E} \psi_A \to \min$	
$\mu_A(x_A) \geq 0$	$f_A(x_A) \leq \psi_A$	$A \subseteq V,\ x_A \in X_A$
$\sum_{x_A} \mu_A(x_A) = 1$	$\psi_A \lesseqgtr 0$	$A \subseteq V$

- Suppose that all possible (i.e., for all $A \subseteq V$) zero constraints are added but marginalisation conditions are omitted. The optimal value of the LP pair is

$$\sum_{A\subseteq V} \max_{x_A} f_A(x_A) = \sum_{A\in E} \max_{x_A} f_A(x_A)$$

Relaxation is given by marginalisation conditions

primal program	**dual program**	
$f\mu \to \max$	$\sum_{A\in E} \psi_A \to \min$	
$\mu_A(x_A) \geq 0$	$f_A^{\varphi}(x_A) \leq \psi_A$	$A \subseteq V,\ x_A \in X_A$
$\sum_{x_A} \mu_A(x_A) = 1$	$\psi_A \lesseqgtr 0$	$A \subseteq V$
$\sum_{x_{A\setminus B}} \mu_A(x_A) = \mu_B(x_B)$	$\varphi_{A,B}(x_B) \lesseqgtr 0$	$(A,B) \in J,\ x_B \in X_B$

- Suppose that all possible (i.e., for all $A \subseteq V$) zero constraints are added but marginalisation conditions are omitted. The optimal value of the LP pair is

$$\sum_{A\subseteq V} \max_{x_A} f_A(x_A) = \sum_{A\in E} \max_{x_A} f_A(x_A)$$

- Couple distribution pairs (μ_A, μ_B) for $(A, B) \in J$ by marginalisation, where

$$J \subseteq I(2^V) = \{\, (A,B) \mid B \subset A \subseteq V \,\} = \text{ the inclusion relation on } 2^V$$

Now, tightness of the relaxation is determined by J alone!
Zero constraints not participating in J are superfluous.

The same graphically...

- The set 2^V of all subsets (except $\emptyset$) of $V = (1, 2, 3, 4)$.

1234

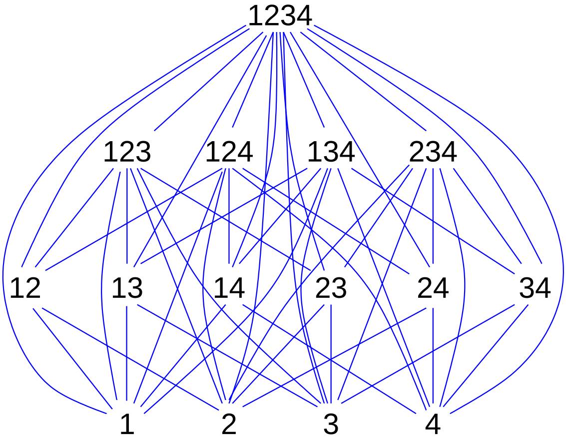
123 124 134 234

12 13 14 23 24 34

1 2 3 4

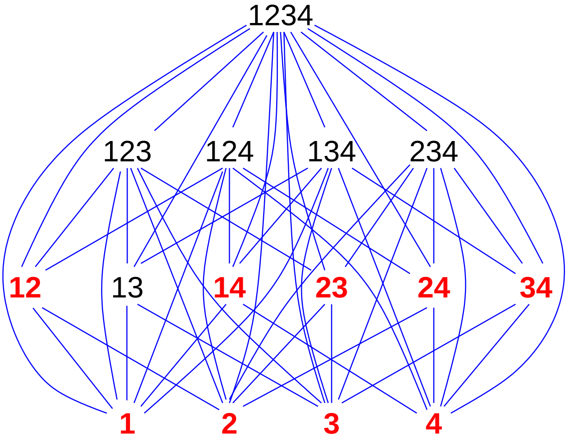
The same graphically...

- The set 2^V of all subsets (except $\emptyset$) of $V = (1, 2, 3, 4)$.
- Inclusion relation $I(2^V)$.

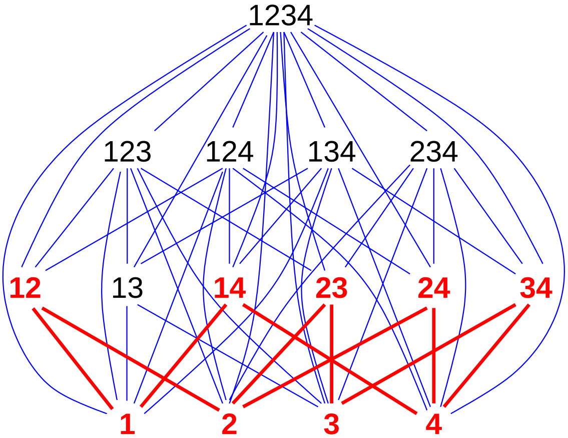


The same graphically...

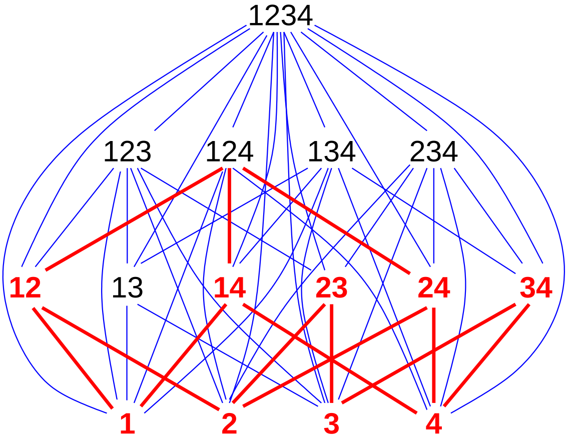
- The set 2^V of all subsets (except $\emptyset$) of $V = (1, 2, 3, 4)$.
- Inclusion relation $I(2^V)$.
- Non-zero constraints, given by hypergraph E.



- The set 2^V of all subsets (except $\emptyset$) of $V = (1, 2, 3, 4)$.
- Inclusion relation $I(2^V)$.
- Non-zero constraints, given by hypergraph E.
- Imposing marginalisations $J = I(E)$ yields the ordinary LP relaxation.



- The set 2^V of all subsets (except $\emptyset$) of $V = (1, 2, 3, 4)$.
- Inclusion relation $I(2^V)$.
- Non-zero constraints, given by hypergraph E.
- Imposing marginalisations $J = I(E)$ yields the ordinary LP relaxation.
- Adding these three pairs (A, B) to J tightens the relaxation.
  This requires adding zero constraint $f_{124} \equiv 0$.



Hierarchy of polyhedral relaxations

$J_1 \subseteq J_2$ implies that relaxation J_1 is not tighter than J_2. Therefore:

Theorem

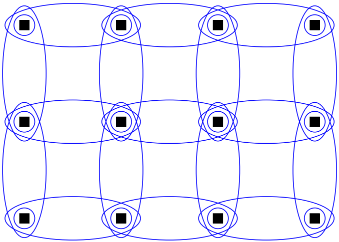
All possible sets $J \subseteq I(2^V)$ form a hierarchy of relaxations, partially ordered by the inclusion relation on $I(2^V)$.

In particular:

$J = \emptyset$	the weakest relaxation (the sum of independent maxima of all the constraints).
$J = I(E)$	the ordinary LP relaxation [Schlesinger-76, other works...]
$J = I(2^V)$	the exact solution (the same as adding the single zero constraint $f_V \equiv 0$)

Example: Adding zero 4-cycle constraints to binary problems

- Randomly draw instances of a binary problem from an instance type.
- Count the number of instances solved exactly by an LP relaxation.
- Two relaxations tested:
  - Plain LP relaxation without zero constraints

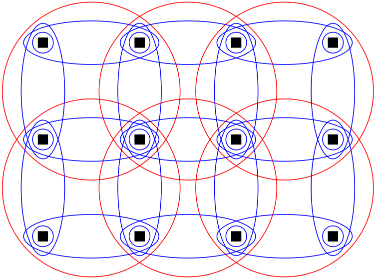


without zero constraints

type	image side	$\lvert X_v \rvert$	r_{plain}
random	15	5	0.01
random	25	3	0.00
random	100	3	0.00
Potts	15	5	0.79
Potts	25	5	0.48
Potts	100	5	0.00
lines	10	4	0.72
lines	25	4	0.00
curve	10	9	0.17
curve	15	9	0.00
curve	25	9	0.00
Pi	15	5	0.00

Example: Adding zero 4-cycle constraints to binary problems

- Randomly draw instances of a binary problem from an instance type.
- Count the number of instances solved exactly by an LP relaxation.
- Two relaxations tested:
  - Plain LP relaxation without zero constraints
  - LP relaxation augmented by 4-ary zero constraints on neighboring variables.



with zero constraints

type	image side	$\|X_v\|$	r_{plain}	r_{4cycle}
random	15	5	0.01	1.00
random	25	3	0.00	0.98
random	100	3	0.00	0.72
Potts	15	5	0.79	0.99
Potts	25	5	0.48	0.98
Potts	100	5	0.00	0.81
lines	10	4	0.72	0.88
lines	25	4	0.00	0.00
curve	10	9	0.17	0.65
curve	15	9	0.00	0.24
curve	25	9	0.00	0.00
Pi	15	5	0.00	0.82

Consider the LP relaxation $\max \left\{ f\mu \;\middle|\; \mu \in P \right\}$ of ILP $\max \left\{ f\mu \;\middle|\; \mu \in P \cap \mathbb{Z}^n \right\}$.

Consider the LP relaxation $\max \{ f\mu \mid \mu \in P \}$ of ILP $\max \{ f\mu \mid \mu \in P \cap \mathbb{Z}^n \}$.

Cutting plane algorithm for general ILP (in primal space)

Cutting plane algorithm

Consider the LP relaxation $\max\left\{ f\mu \;\middle|\; \mu \in P \right\}$ of ILP $\max\left\{ f\mu \;\middle|\; \mu \in P \cap \mathbb{Z}^n \right\}$.

Cutting plane algorithm for general ILP (in primal space)

1: $P' \leftarrow P$

Cutting plane algorithm

Consider the LP relaxation $\max \{ f\mu \mid \mu \in P \}$ of ILP $\max \{ f\mu \mid \mu \in P \cap \mathbb{Z}^n \}$.

Cutting plane algorithm for general ILP (in primal space)

$P' \leftarrow P$
loop
  Find a maximiser μ^* of $\max \{ f\mu \mid \mu \in P' \}$.
end loop

Consider the LP relaxation $\max\left\{ f\mu \;\middle|\; \mu \in P \right\}$ of ILP $\max\left\{ f\mu \;\middle|\; \mu \in P \cap \mathbb{Z}^n \right\}$.

Cutting plane algorithm for general ILP (in primal space)

$P' \leftarrow P$
loop
    Find a maximiser μ^* of $\max\left\{ f\mu \;\middle|\; \mu \in P' \right\}$.
    Find a half-space H such that $\mu^* \notin H \supseteq P \cap \mathbb{Z}^n$. If none exists, halt.

end loop

Cutting plane algorithm

Consider the LP relaxation $\max\left\{ f\mu \mid \mu \in P \right\}$ of ILP $\max\left\{ f\mu \mid \mu \in P \cap \mathbb{Z}^n \right\}$.

Cutting plane algorithm for general ILP (in primal space)

$P' \leftarrow P$
loop
    Find a maximiser μ^* of $\max\left\{ f\mu \mid \mu \in P' \right\}$.
    Find a half-space H such that $\mu^* \notin H \supseteq P \cap \mathbb{Z}^n$. If none exists, halt.
    $P' \leftarrow P' \cap H$
end loop

Consider the LP relaxation $\max\left\{ f\mu \,\middle|\, \mu \in P \right\}$ of ILP $\max\left\{ f\mu \,\middle|\, \mu \in P \cap \mathbb{Z}^n \right\}$.

Cutting plane algorithm for general ILP (in primal space)

Cutting plane algorithm for WCSP (in dual space)

Cutting plane algorithm

Consider the LP relaxation $\max\{\, f\mu \mid \mu \in P \,\}$ of ILP $\max\{\, f\mu \mid \mu \in P \cap \mathbb{Z}^n \,\}$.

Cutting plane algorithm for general ILP (in primal space)

$P' \leftarrow P$
loop
    Find a maximiser μ^* of $\max\{\, f\mu \mid \mu \in P' \,\}$.
end loop

Cutting plane algorithm for WCSP (in dual space)

loop
    Find the equivalent network with the least upper bound.
end loop

Cutting plane algorithm

Consider the LP relaxation $\max\{\, f\mu \mid \mu \in P \,\}$ of ILP $\max\{\, f\mu \mid \mu \in P \cap \mathbb{Z}^n \,\}$.

Cutting plane algorithm for general ILP (in primal space)

$P' \leftarrow P$
loop
    Find a maximiser μ^* of $\max\{\, f\mu \mid \mu \in P' \,\}$.
    Find a half-space H such that $\mu^* \notin H \supseteq P \cap \mathbb{Z}^n$. If none exists, halt.

end loop

Cutting plane algorithm for WCSP (in dual space)

loop
    Find the equivalent network with the least upper bound.
    Find $A \notin E$ such that the CSP formed by active joint states restricted on hypergraph $E \cap 2^A$ is unsatisfiable. If none exists, halt.

end loop

Cutting plane algorithm

Consider the LP relaxation $\max\{\, f\mu \mid \mu \in P \,\}$ of ILP $\max\{\, f\mu \mid \mu \in P \cap \mathbb{Z}^n \,\}$.

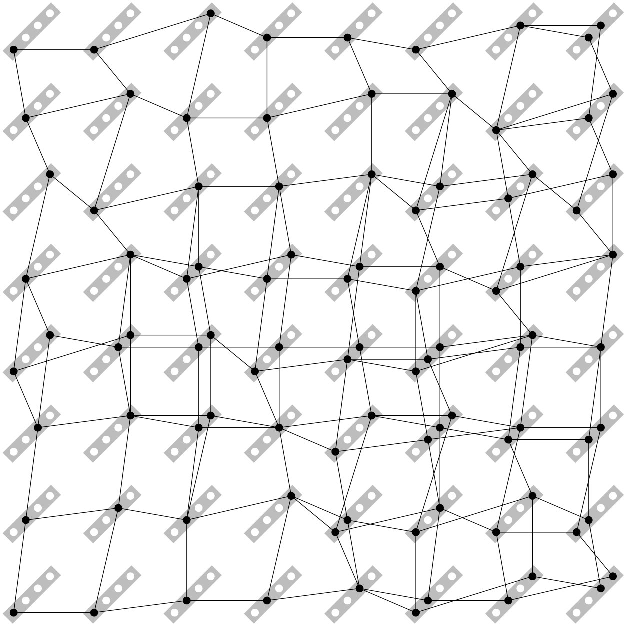
Cutting plane algorithm for general ILP (in primal space)

```
P' ← P
loop
    Find a maximiser μ* of max{ fμ | μ ∈ P' }.
    Find a half-space H such that μ* ∉ H ⊇ P ∩ ℤⁿ. If none exists, halt.
    P' ← P' ∩ H
end loop
```

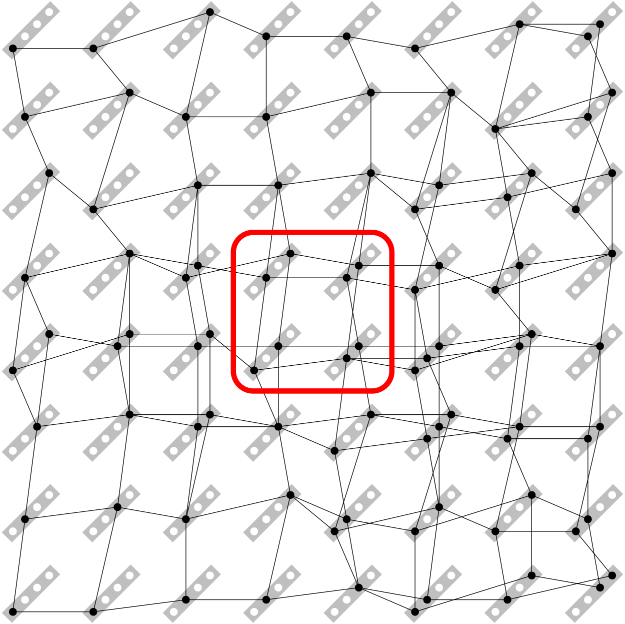
Cutting plane algorithm for WCSP (in dual space)

```
loop
    Find the equivalent network with the least upper bound.
    Find A ∉ E such that the CSP formed by active joint states restricted on
    hypergraph E ∩ 2^A is unsatisfiable. If none exists, halt.
    f_A ← 0;  E ← E ∪ {A}
end loop
```

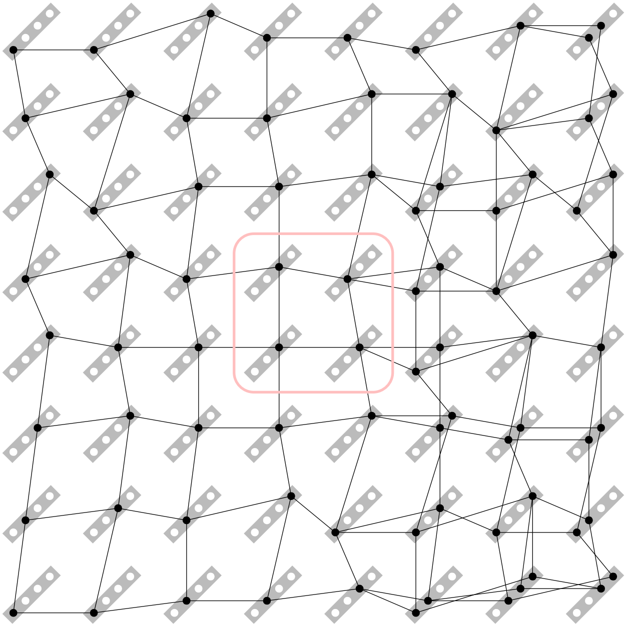
Example run of cutting plane algorithm

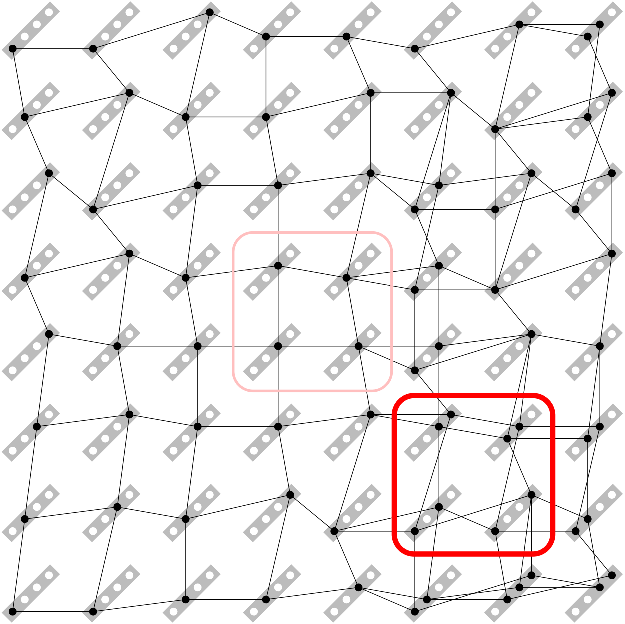


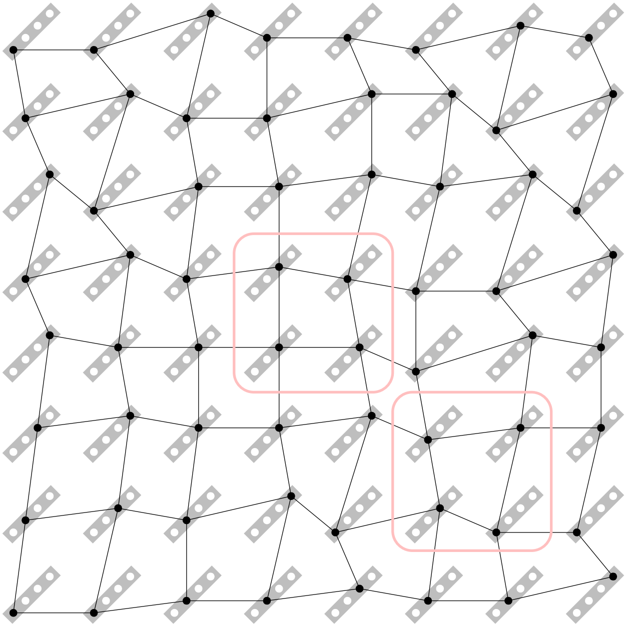
Example run of cutting plane algorithm



Example run of cutting plane algorithm







Integral hull of LP relaxation

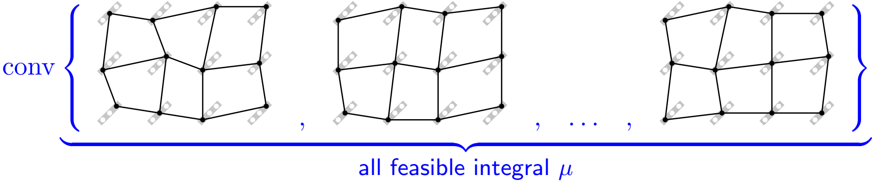
LP relaxation is a common approach to combinatorial optimisation problems:

$$\max\{\, f\mu \mid \mu \in \underbrace{\mathrm{conv}(P \cap \mathbb{Z}^n)}_{\substack{\text{integral hull}\\ \text{of } P}} \,\} = \overbrace{\max\{\, f\mu \mid \mu \in P \cap \mathbb{Z}^n \,\}}^{\text{integer LP}} \leq \overbrace{\max\{\, f\mu \mid \mu \in P \,\}}^{\text{LP relaxation}}$$

- P is a convex polyhedron with a tractable number of facets.
- For NP-hard ILPs, $\mathrm{conv}(P \cap \mathbb{Z}^n)$ has much more facets than P and they cannot be described in a 'short' way. Facets of $\mathrm{conv}(P \cap \mathbb{Z}^n)$ that are not facets of P are good cutting planes.
- All vertices of $\mathrm{conv}(P \cap \mathbb{Z}^n)$ are integral while P may have also fractional vertices.

Integral hull of WCSP = marginal polytope

Integral hull of WCSP is the set



Definition (Marginal polytope [Wainwright-etal-2003])

Marginal polytope (associated with hypergraph $E \subseteq 2^V$ and domains $\{X_v\}$) is a set of mappings $\mu\colon T(E) \to [0,1]$ defined as follows: μ belongs to the marginal polytope iff there exists a function $\mu_V\colon X_V \to [0,1]$ such that $\sum_{x_V} \mu_V(x_V) = 1$ and

$$\forall A \in E,\, x_A \in X_A : \qquad \sum_{x_{V \setminus A}} \mu_V(x_V) = \mu_A(x_A)$$

That is, μ is a set of valid marginals of some global distribution μ_V.

Theorem

Marginal polytope is the integral hull of WCSP.